TENSOR ANALYSIS OF ANOVA DECOMPOSITION

TECHNICAL REPORT NO. 6

AKIMICHI TAKEMURA

NOVEMBER 1982

DEPARTMENT OF STATISTICS

STANFORD   UNIVERSITY

STANFORD, CALIFORNIA

TENSOR ANALYSIS OF ANOVA DECOMPOSITION

TECHNICAL REPORT NO. 6

AKIMICHI TAKEMURA
STANFORD UNIVERSITY

NOVEMBER 1982

U. S. ARMY RESEARCH OFFICE

CONTRACT DAAG29-82-K-0156

DEPARTMENT OF STATISTICS
STANFORD   UNIVERSITY
STANFORD, CALIFORNIA

§1 Introduction.

The purpose of this paper is to demonstrate the almost complete analogy between ANOVA for general n-factor crossed layouts and the ANOVA type decomposition of square integrable statistics used in the literature on U-statistics and in connection with the Jackknife estimate of variance (Efron and Stein(1981), Bhargava(1980), Karlin and Rinott(1982)). This will be done using notions and notations of tensor analysis and multilinear algebra. It will be clear that the latter is just an infinite dimensional generalization of the former. Usually the analogy between the two is understood in an operational sense, namely how the higher order interaction terms are defined by an inclusion-exclusion argument. Various subclass means are added and subtracted in the usual ANOVA; various conditional expectations in the second case. Then the orthogonality of the interaction terms are proved. For ANOVA Mann(1949, Chapter 5) gives a classical treatment. See also Han(1977) for a treatment in modern terminology. By using tensor analysis we can in a sense reverse the argument. Orthogonal subspaces of an appropriate vector space can be directly described. Only the dimensionality is different in the two cases. The inclusion-exclusion pattern then follows from the form of the orthogonal projectors onto these subspaces.

In ANOVA and experimental design tensor approach has been employed by a number of people. It provides a natural and powerful tool for treating general n-factor crossed layouts and other designs. Unfortunately terminology and notation were not standardized, in particular an essentially same notion has been called *tensor, Kronecker, direct,* or *outer product*. Approaches employed were sometimes elementary, sometimes more abstract. This is one of the reasons why this approach has not been very often taught.

In the field of experimental design, Kurkjian and Zelen(1962) introduced a "calculus for factorial arrangements". Following this work there have been many papers using direct product notation for construction and analysis of various designs, including Kurkjian and Zelen(1963), Zelen and Federer(1964,1965), Federer and Zelen(1966), Bock(1963), Paik and Federer(1974), Cotter, John, and Smith(1973), Cotter (1974,1975), John and Dean(1975a,b). The terminology and notational conventions introduced by Kurkjian and Zelen(1962) seem to be rather arbitrary. Connection between their "calculus" and the standard tensor

analysis or multilinear algebra was not made clear. Another drawback is that they confined their theory to the usual matrix theory and multilinear aspects tend to be lost. For example they define direct product of matrices as a partitioned matrix of a larger dimensionality (this is still a common practice today in statistics). But this introduces an unpleasant ordering of indices and the symmetry inherent in the problem becomes obscured.

Another group of people employing this technique are found in the coordinate-free approach in linear models, for example Jacobsen(1968), Eaton(1970), Haberman(1975). Jacobsen(1968) seems to be the first systematic treatment of ANOVA from the viewpoint of multilinear algebra. In addition to the new viewpoint his treatment of the nested model and the missing observation method is interesting. Unfortunately his results do not seem to have been published in a more widely available form and has been almost forgotten in the later literature. Furthermore his treatment suffers from excessive mathematical formalism and arbitrary notational conventions. Later Haberman(1975) gave a thorough treatment which can be regarded as a standard reference so far. One problem with these mathematical treatments is that an essentially elementary nature of the approach and practical computational aspects are often difficult to grasp.

In Section 2 we define tensors as multidimensional arrays as in the usual tensor analysis (Sokolnikoff(1964), Chapter 2). By doing this the unpleasant ordering of indices mentioned above is avoided. Operations on these arrays are explicitly described. In any high level computer language multidimensional arrays can be used as easily as matrices, so this approach can be immediately incorporated in computer programs. Standard terminology of tensor analysis and multilinear algebra will be employed. Furthermore we develop the theory in such a way that it can be easily generalized to L^2-spaces.

In Section 3 we briefly look at the general n-factor crossed layout.

In Section 4 we treat the ANOVA type decomposition of a statistic with finite second moment by generalizing the results of Section 2 and 3 to L^2-spaces. The decomposition was first introduced by Hoeffding(1948) in connection with U-statistics. Often the linear terms of this decomposition (corresponding to the main effects in ANOVA) are called Hajek projection following Hajek(1968) and used extensively to prove asymptotic normality of various statistics. See Serfling(1980) for further references. Recently more

attention is paid to the full decomposition. Rubin and Vitale(1980) developed a general asymptotic theory of U-statistics using the full decomposition. Efron and Stein(1981) used the decomposition in their study of the Jackknife estimate of variance. Further results and generalizations are given in Bhargava(1980) and Karlin and Rinott(1982). In the absense of a standard reference for the decomposition each of them gave a definition of the decomposition using various variations of an inclusion-exclusion argument. Our approach is different from these as mentioned at the beginning.

§2 Tensor products of vectors, matrices, vector spaces and subspaces.

In this section we develop a theory of tensors. Particular references used in this section are Greub(1978), Chapter 1 and Sokolnikoff(1964), Chapter 2. A full abstract treatment can be found in Chapter 1 of Greub(1978). In a pure mathmatical treatment tensors are developed in a coordinate-free way (Greub(1978)). This is elegant but not desirable from the viewpoint of computational applicability in statistics. On the other hand the traditional tensor analysis (Sokolnikoff(1964)) is more practical but is too closely tied to physics and much emphasis is placed on curvilinear coordinates which we do not need here. We take appropriate notions and notations needed from both of them. Proofs can be found in various references given above and hence omitted below except for a few places.

Let R^m be the set of all column vectors $x = (x^1, \ldots, x^m)'$ with m elements of real numbers. To denote the components of a column (or *contravariant*) vector we use superscripts following the traditional notation in tensor analysis. Vector addition and scalar multiplication are defined in the usual componentwise way. Now *tensor (Kronecker, direct, outer) product* $x \otimes y$ of x ($\in R^m$) and y ($\in R^n$) is a two-dimensional array defined by a componentwise multiplication:

$$(2.1) \qquad (x \otimes y)^{ij} = x^i \cdot y^j.$$

Namely, $x \otimes y$ is a two-dimensional array of dimensions m and n whose (i, j) element is $x^i y^j$ ($i = 1, \ldots, m$, $j = 1, \ldots, n$). Now we define an addition of tensor products in a componentwise way.

$$(2.2) \qquad (ax \otimes y + b\tilde{x} \otimes \tilde{y})^{ij} = ax^i y^j + b\tilde{x}^i \tilde{y}^j,$$

where a, b are scalars. This leads to a vector space generated by $\{\, x \otimes y, \quad x \in R^m, \quad y \in R^n \,\}$ which we denote by $R^m \otimes R^n$. Namely

$$
\begin{aligned}
(2.3) \qquad R^m \otimes R^n &= \operatorname{span}\{\, x \otimes y, \quad x \in R^m, \quad y \in R^n \,\} \\
&= \{\, \sum_{i=1}^{k} a_i (\underset{i}{x} \otimes \underset{i}{y}), \quad k : finite \,\}.
\end{aligned}
$$

Here the index i is written directly below the corresponding vectors x and y because usual subscripts are used as covariant indices in tensor analysis. This point will be discussed later in this section in connection with linear transformations. $R^m \otimes R^n$ is called the tensor product of R^m and R^n. A general element $u \in R^m \otimes R^n$ is called simply as a tensor.

As one might expect, $R^m \otimes R^n$ is just the set of all two-dimensional arrays of dimensions m and n. We will make this point clear in a couple of propositions.

Lemma 2.1. $x \otimes y$ *is bilinear in x and y. Namely*

$$
\begin{aligned}
(2.4) \qquad (ax + b\tilde{x}) \otimes y &= a(x \otimes y) + b(\tilde{x} \otimes y), \\
x \otimes (cy + d\tilde{y}) &= c(x \otimes y) + d(x \otimes \tilde{y}),
\end{aligned}
$$

where a, b, c, d are scalars.

Let $\underset{i}{\overset{m}{e}}$ denote a vector in R^m whose i-th element is one and other elements are zero. $\{\underset{i}{\overset{m}{e}}, i = 1, \ldots, m\}$ forms an obvious basis of R^m. Now consider $\underset{i}{\overset{m}{e}} \otimes \underset{j}{\overset{n}{e}}$ which has 1 in (i, j)-position and 0 everywhere else. Then

Proposition 2.1. $\{\underset{i}{\overset{m}{e}} \otimes \underset{j}{\overset{n}{e}}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n\}$ *is a basis of $R^m \otimes R^n$.*

hence

Corollary 2.1. $\dim(R^m \otimes R^n) = mn$ *and $R^m \otimes R^n$ coincides with the set of all two-dimensional arrays of dimensions m and n.*

Remark 2.1. An element $u \in R^m \otimes R^n$ which can be written as $u = x \otimes y$ for some $x \in R^m$, $y \in R^n$ is called *decomposable*. $R^m \otimes R^n$ does not consist only of decomposable elements. This is easily seen by noting that $x \otimes y$ is of "rank 1" in the terminology of the usual matrix theory.

In R^m we have the usual inner product. In $R^m \otimes R^n$ a natural inner product is defined in an analogous way. Let $u, v \in R^m \otimes R^n$. Then we define

$$(2.5) \qquad (u, v) = \sum_{i=1}^{m} \sum_{j=1}^{n} u^{ij} v^{ij}.$$

Proposition 2.2. *For two decomposable elements* $x \otimes \tilde{x}, y \otimes \tilde{y}$ *of* $R^m \otimes R^n$ *we have*

$$(2.6) \qquad (x \otimes \tilde{x}, y \otimes \tilde{y}) = (x, y) \cdot (\tilde{x}, \tilde{y}).$$

Proof:

$$\begin{aligned}
(x \otimes \tilde{x}, y \otimes \tilde{y}) &= \sum_{i,j} (x \otimes \tilde{x})^{ij} \cdot (y \otimes \tilde{y})^{ij} \\
&= \sum_{i,j} x^i \tilde{x}^j y^i \tilde{y}^j \\
&= \sum_{i} x^i y^i \sum_{j} \tilde{x}^j \tilde{y}^j \\
&= (x, y) \cdot (\tilde{x}, \tilde{y}).
\end{aligned}$$

Now we proceed to define tensor products of more than two vectors. Let $\underset{i}{x} \in R^{m_i}, i = 1, \ldots, k$. Then $\underset{1}{x} \otimes \cdots \otimes \underset{k}{x}$ is defined to be a k-dimensional array of dimensions $m_1, \ldots, m_k$ such that

$$(2.7) \qquad (\underset{1}{x} \otimes \cdots \otimes \underset{k}{x})^{i_1 \cdots i_k} = \underset{1}{x}^{i_1} \cdots \underset{k}{x}^{i_k}.$$

Addition is defined componentwise and the space generated by $\{ \underset{1}{x} \otimes \cdots \otimes \underset{k}{x}, \ \underset{i}{x} \in R^{m_i}, \ i = 1, \ldots, k \}$ is called the tensor product of $R^{m_1}, \ldots, R^{m_k}$ and denoted by $R^{m_1} \otimes \cdots \otimes R^{m_k}$ or $\otimes_{i=1}^{k} R^{m_i}$. Lemma 2.1, Proposition 2.1, Corollary 2.1 hold for $k > 2$ with obvious modifications. Now for general element u, v of $\otimes_{i=1}^{k} R^{m_i}$ we define the natural inner product by

$$(2.8) \qquad (u, v) = \sum_{i_1=1}^{m_1} \cdots \sum_{i_k=1}^{m_k} u^{i_1 \ldots i_k} v^{i_1 \ldots i_k}.$$

Then analogous to (2.6) for two decomposable elements $\underset{1}{x} \otimes \cdots \otimes \underset{k}{x}, \underset{1}{y} \otimes \cdots \otimes \underset{k}{y}$ of $\otimes_{i=1}^{k} R^{m_i}$ we have

$$(2.9) \qquad (\underset{1}{x} \otimes \cdots \otimes \underset{k}{x}, \underset{1}{y} \otimes \cdots \otimes \underset{k}{y}) = (\underset{1}{x}, \underset{1}{y}) \cdot (\underset{2}{x}, \underset{2}{y}) \cdots (\underset{k}{x}, \underset{k}{y}).$$

Remark 2.2. If x_i and y_i are orthogonal for some i, then $\underset{1}{x}\otimes\cdots\otimes\underset{k}{x}$ and $\underset{1}{y}\otimes\cdots\otimes\underset{k}{y}$ are orthogonal.

Next we consider subspaces and its orthogonal complements. Let $U_1,\ldots,U_k$ be subspaces of $R^{m_1},\ldots,R^{m_k}$ respectively. Then a subspace $U_1\otimes\cdots\otimes U_k$ of $R^{m_1}\otimes\cdots\otimes R^{m_k}$ is defined to be the subspace generated by $\{\,\underset{1}{x}\otimes\cdots\otimes\underset{k}{x},\ \underset{i}{x}\in U_i,\ i=1,\ldots,k\,\}$. Namely

$$(2.10)\qquad U_1\otimes\cdots\otimes U_k = \text{span}\{\underset{1}{x}\otimes\cdots\otimes\underset{k}{x},\quad \underset{i}{x}\in U_i,\quad i=1,\ldots,k\,\}.$$

Let $U_i^{\perp}$ denote the orthogonal complement of U_i in R^{m_i}. For convenience we define $U_i^0 = U_i, U_i^1 = U_i^{\perp}$. Then we have

Theorem 2.1 2^k *subspaces* $\{\otimes_{i=1}^{k}U_i^{\epsilon_i},\quad \epsilon_i = 0,1,\quad i=1,\ldots,k\,\}$ *form a decomposition of* $\otimes_{i=1}^{k}R^{m_i}$ *into mutually orthogonal subspaces.*

This is clear by taking appropriate orthonormal basis of R^{m_i}, $i=1,\ldots,k$ and applying Remark 2.2.

Corollary 2.2.

$$(2.11)\qquad (U_1\otimes\cdots\otimes U_k)^{\perp} = \text{span}\{\otimes_{i=1}^{k}U_i^{\epsilon_i},\ \epsilon_i = 1\ for\ some\ i\,\}.$$

Now we are going to define tensor product of matrices. An $n\times m$ matrix A is considered to represent a linear transformation from R^m to R^n. In this sense we want to distinguish matrices from two-dimensional arrays (elements of $R^n\otimes R^m$). In tensor analysis this is done by writing the second index as subscripts. Namely (i,j) element of a matrix A is denoted by A_j^i. Superscripts are called *contravariant indices* and subscripts as called *covariant indices*. The reason behind this is discussed in Remark 2.4 below. Now let $\underset{i}{A}$ be $n_i\times m_i$ matrices, $i=1,\ldots,k$. We want to define a tensor product of $\underset{1}{A},\ldots,\underset{k}{A}$ in a meaningful way. For notational convenience we first do this for the case $k=2$. For matrices A $(n_1\times m_1)$ and B $(n_2\times m_2)$ we define $A\otimes B$ as a four-dimensional array with two contravariant indicies i_1,i_2 and two covariant indices j_1,j_2 such that

$$(2.12)\qquad (A\otimes B)_{j_1 j_2}^{i_1 i_2} = A_{j_1}^{i_1}\cdot B_{j_2}^{i_2}.$$

This is again a componentwise multiplication as in (2.1). Now $A\otimes B$ defines a linear transformation from $R^{m_1}\otimes R^{m_2}$ to $R^{n_1}\otimes R^{n_2}$ as follows. Let $u\in R^{m_1}\otimes R^{m_2}$ then $v=$

$(A \otimes B)u \in R^{n_1} \otimes R^{n_2}$ is defined by

(2.13)
$$v^{i_1 i_2} = \sum_{j_1=1}^{m_1} \sum_{j_2=1}^{m_2} (A \otimes B)^{i_1 i_2}_{j_1 j_2} u^{j_1 j_2}$$
$$= \sum_{j_1=1}^{m_1} \sum_{j_2=1}^{m_2} A^{i_1}_{j_1} \cdot B^{i_2}_{j_2} \cdot u^{j_1 j_2}.$$

As in (2.6) we have

Proposition 2.3.   *For a decomposable element $x \otimes y$ of $R^{m_1} \otimes R^{m_2}$*

(2.14) $$(A \otimes B)(x \otimes y) = Ax \otimes By.$$

Remark 2.3.   We could have used (2.14) as a definition of $A \otimes B$. It shows how $A \otimes B$ maps decomposable elements of $R^{m_1} \otimes R^{m_2}$. Since the decomposable elements generate $R^{m_1} \otimes R^{m_2}$, $A \otimes B$ for general elements can be defined by linearity. This is more elegant mathematically but for practical applications formula (2.13) will be useful. The same remark applies to Proposition 2.2.

Generalization of the above argument to tensor product of more than 2 matrices is immediate. Instead of (2.13) and (2.14) we have

(2.15) $$(\underset{1}{A} \otimes \cdots \otimes \underset{k}{A}) u^{i_1 \cdots i_k} = \sum_{j_1, \ldots, j_k} \underset{1}{A}^{i_1}_{j_1} \cdots \underset{k}{A}^{i_k}_{j_k} u^{j_1 \cdots j_k},$$

(2.16) $$(\underset{1}{A} \otimes \cdots \otimes \underset{k}{A})(\underset{1}{x} \otimes \cdots \otimes \underset{k}{x}) = \underset{1}{A}\underset{1}{x} \otimes \cdots \otimes \underset{k}{A}\underset{k}{x}.$$

respectively.

Remark 2.4. R^m was defined as the set of column or contravariant vectors. The dual space R^{m*} can be defined as the set of row or covariant vectors whose components are denoted with subscripts $x = (x_1, \ldots, x_m)$. Then $R^{m*} \otimes R^{n*}$, $R^m \otimes R^{n*}$, etc., can be defined in a similar way as $R^m \otimes R^n$ is defined by (2.1)–(2.3). By the natural isomorphism between $R^m \otimes R^{n*}$ and the space of all linear transformations from R^n to R^m (see Greub(1978), Section 1.28) a linear transformation A can be identified with an element of $R^m \otimes R^{n*}$ and has one contravariant index and one covariant index.

Our last item in this section is a discussion on orthogonal projectors. Let V be a vector space and $U \subset V$ be a subspace. A linear transformation P_U from V to itself is called the orthogonal projector onto U if

(2.17)
$$P_U x = x \quad for \quad x \in U,$$
$$P_U x = 0 \quad for \quad x \in U^\perp.$$

Theorem 2.2. *Let* $U_i \subset R^{m_i}$, $i = 1, \ldots, n$ *be subspaces and* P_{U_i} *be the orthogonal projectors onto* U_i *in* $R^{m_i}, i = 1, \ldots, n$. *Then the orthogonal projector onto* $\otimes_{i=1}^n U_i \subset \otimes_{i=1}^k R^{m_i}$ *is given by* $P_{U_1} \otimes \cdots \otimes P_{U_n}$.

Proof: Let $\underset{i}{x} \in U_i$, $i = 1, \ldots, k$. Then by (2.16)

(2.18)
$$(P_{U_1} \otimes \cdots \otimes P_{U_k})(\underset{1}{x} \otimes \cdots \otimes \underset{k}{x}) = P_{U_1}\underset{1}{x} \otimes \cdots \otimes P_{U_k}\underset{k}{x}$$
$$= \underset{1}{x} \otimes \cdots \otimes \underset{k}{x}.$$

Hence for general elements u of $U_1 \otimes \cdots \otimes U_k$ we have $(P_{U_1} \otimes \cdots \otimes P_{U_k})u = u$ by linearity. Now by Corollary 2.2 $(U_1 \otimes \cdots \otimes U_k)^\perp$ is generated by $\{ \underset{1}{x} \otimes \cdots \otimes \underset{k}{x}, \underset{i}{x} \in U_i^\perp \ for \ some \ i \}$. For such $\underset{1}{x} \otimes \cdots \otimes \underset{k}{x}$

(2.19)
$$(P_{U_1} \otimes \cdots \otimes P_{U_k})(\underset{1}{x} \otimes \cdots \otimes \underset{k}{x}) = P_{U_1}\underset{1}{x} \otimes \cdots \otimes P_{U_k}\underset{k}{x}$$
$$= 0.$$

Hence by linearity

$$(P_{U_1} \otimes \cdots \otimes P_{U_k})(U_1 \otimes \cdots \otimes U_k)^\perp = \{ 0 \}.$$

∎

See Haberman(1975), Lemma 8, for an alternative proof using the fact that $P_{U_1} \otimes \cdots \otimes P_{U_k}$ is idempotent and self-adjoint.

§3 ANOVA for crossed layouts.

Now we take a brief look at ANOVA for an n-factor crossed layout with single observation per cell. For more detailed treatments of various designs see the references

given in Section 1. For each combination of factor levels $(i_1, \ldots, i_n)$ we have an observation $x^{i_1 \cdots i_n}$. Therefore the set of observartions $\{x^{i_1 \cdots i_n}\}$ can be considered as a (random) tensor $x \in \bigotimes_{i=1}^{n} R^{m_i}$ ANOVA is essentially a decomposition of $\bigotimes_{i=1}^{n} R^{m_i}$ into mutually orthogonal subspaces. When all interactions are considered it is decomposed into 2^n subspaces. Usually this is done by an inclusion-exclusion argument. Here we give the desired decomposition directly as follows. Let $\overset{m_i}{1}$ be a vector in R^{m_i} with all components equal to 1. Let $U_i = \mathrm{span}\{\overset{m_i}{1}\}$ and consider the decomposition of $\bigotimes_{i=1}^{n} R^{m_i}$ in Theorem 2.1. We use the notational convention of Theorem 2.1. Following Scheffé(1959), Section 4.6 let $\mathcal{L}_{i_1 \ldots i_k}$ denote the $(i_1, \ldots, i_k)$ -interaction subspace for $1 \le i_1 < \ldots < i_k \le n$, $k = 0, \ldots, n$. We claim that

$$(3.1) \qquad \mathcal{L}_{i_1 \ldots i_k} = U_1^{\epsilon_1} \otimes \cdots \otimes U_n^{\epsilon_n},$$

where

$$\begin{aligned} \epsilon_i &= 1 \qquad \textit{if } i \in \{i_1, \ldots, i_k\}, \\ &= 0 \qquad \textit{otherwise.} \end{aligned}$$

This can be shown by considering the orthogonal projector onto the right hand side of (3.1). Note that the orthogonal projector onto $U_i = \mathrm{span}\{\overset{m_i}{1}\}$ is given by (in matrix form)

$$(3.2) \qquad \underset{i}{F} = \frac{1}{m_i} \overset{m_i}{1}\overset{m_i}{1}{}'.$$

For $x = (x^1, \ldots, x^{m_i})'$ we have $\underset{i}{F}x = (\bar{x}, \ldots, \bar{x})'$. Furthermore $\overset{m_i}{I} - \underset{i}{F}$ is the orthogonal projector onto $U_i^{\perp}$, where $\overset{m_i}{I}$ denotes the $m_i \times m_i$ identity matrix. Now by Theorem 2.2 the orthogonal projector onto the right hand side of (3.1) is given by

$$(3.3) \qquad P_{i_1 \ldots i_k} = \underset{1}{Q} \otimes \cdots \otimes \underset{k}{Q},$$

where

$$\begin{aligned} \underset{i}{Q} &= \overset{m_i}{I} - \underset{i}{F} \qquad \textit{if } i \in \{i_1, \ldots, i_k\}, \\ &= \underset{i}{F} \qquad \textit{otherwise.} \end{aligned}$$

For example consider P_{12}:

$$
\begin{aligned}
P_{12} &= (\overset{m_1}{I} - \underset{1}{F})\otimes(\overset{m_2}{I} - \underset{2}{F})\otimes\underset{3}{F}\otimes\cdots\otimes\underset{n}{F} \\
&= \overset{m_1}{I}\otimes\overset{m_2}{I}\otimes\underset{3}{F}\otimes\cdots\otimes\underset{n}{F} - \underset{1}{F}\otimes\overset{m_2}{I}\otimes\underset{3}{F}\otimes\cdots\otimes\underset{n}{F} \\
&\quad - \overset{m_1}{I}\otimes\underset{2}{F}\otimes\underset{3}{F}\otimes\cdots\otimes\underset{n}{F} + \underset{1}{F}\otimes\underset{2}{F}\otimes\underset{3}{F}\otimes\cdots\otimes\underset{n}{F}.
\end{aligned}
$$

(3.4)

Operating P_{12} to x we obtain the usual expression:

(3.5)
$$
(P_{12}x)^{i_1\cdots i_n} = \bar{x}^{i_1 i_2\cdots} - \bar{x}^{\cdot i_2\cdots} - \bar{x}^{i_1\cdots} + \bar{x}^{\cdots}.
$$

Note that the expansion of the expression for the projector leads to the inclusion-exclusion. This pattern should be clear for general $P_{i_1\ldots i_k}$. This proves (3.1).

The sum of squares due to $(i_1,\ldots,i_k)$- interaction for an observed tensor x is given by

(3.6)
$$
S_{i_1\ldots i_k} = (P_{i_1\ldots i_k}x, P_{i_1\ldots i_k}x)
$$

Note that the actual computation of (3.6) can be done using (3.3),(2.15) and (2.8).

The degrees of freedom (d.f.) of $(i_1,\ldots,i_k)$-interaction is given by $\dim(\mathcal{L}_{i_1\ldots i_k})$. Noting that $\dim(U_i) = 1$ and $\dim(U_i^\perp) = m_i - 1$ we obtain by Corollary 2.1

(3.7)
$$
d.f.\ of\ (i_1,\ldots,i_k) - interaction = \prod_{j=1}^{k}(m_{i_j} - 1).
$$

§4 ANOVA decomposition of a statistic.

In this section we study the ANOVA type decomposition of a square integrable statistic $S(x_1,\ldots,x_k)$. For this purpose we extend the results in the previous sections to L^2-spaces. Particular references used here are Maurin(1967), Section 3.10 and Murray and von Neumann(1936), Chapter 2. Let $(X_1,\mu_1),\ldots,(X_n,\mu_n)$ be probability spaces. We consider the L^2-space of the product probability space $(X_{i_1},\mu_{i_1})\times\cdots\times(X_{i_k},\mu_{i_k})$:

(4.1)
$$
L^2(X_{i_1},\ldots,X_{i_k}) = \{\phi(x_{i_1},\ldots,x_{i_k})\mid\int\phi^2\mu_{i_1}(dx_{i_1})\cdots\mu_{i_k}(dx_{i_k}) < \infty\}.
$$

Note that $L^2(X_{i_1}, \ldots, X_{i_k}) \subset L^2(X_{j_1}, \ldots, X_{j_\ell})$ if $\{i_1, \ldots, i_k\} \subset \{j_1, \ldots, j_\ell\}$. For simplicity we assume that $X_1, \ldots, X_n$ are locally compact, separable, metrizable spaces so that L^2-spaces in (4.1) are separable. See Dieudonné (1976), Chap. 13.

For notational convenience let n=2, general case being an obvious modification. Let $\phi(x_1) \in L^2(X_1)$, $\psi(x_2) \in L^2(X_2)$. Intuitively we can think of ϕ, ψ as having continuous indices x_1, x_2. Sum of squares is replaced by squared integrals. Now define $\phi \otimes \psi$ by a componentwise multiplication:

$$(4.2) \qquad (\phi \otimes \psi)(x_1, x_2) = \phi(x_1) \cdot \psi(x_2) \in L^2(X_1, X_2).$$

Note that

$$(4.3) \qquad \begin{aligned} &\int (\phi(x_1)\psi(x_2))^2 \mu_1(dx_1)\mu_2(dx_2) \\ &= \int \phi(x_1)^2 \mu_1(dx_1) \int \psi(x_2)^2 \mu_2(dx_2) < \infty. \end{aligned}$$

Hence $\phi \otimes \psi \in L^2(X_1, X_2)$.

Now let $L^2(X_1) \otimes L^2(X_2) \subset L^2(X_1, X_2)$ be defined as in (2.3), namely

$$(4.4) \qquad \begin{aligned} L^2(X_1) \otimes L^2(X_2) &= \mathrm{span}\{ \phi \otimes \psi, \quad \phi \in L^2(X_1), \quad \psi \in L^2(X_2) \} \\ &= closure\, of\, \{ \sum_{i=1}^{k} a_i \phi_i \otimes \psi_i, \quad k : finite \}. \end{aligned}$$

Proposition 4.1. $\quad L^2(X_1) \otimes L^2(X_2) = L^2(X_1, X_2).$

This is a standard construction (Maurin(1967), Example of Section 3.10) and a simple consequence of the following well-known result.

Lemma 4.1. *Let* $\{ \phi_1, \phi_2, \ldots \}$ *and* $\{ \psi_1, \psi_2, \ldots \}$ *be complete orthonormal systems of* $L^2(X_1), L^2(X_2)$ *respectively. Then* $\{ \phi_i \psi_j, \quad i = 1, 2, \ldots, j = 1, 2, \ldots \}$ *is a complete orthonormal system of* $L^2(X_1, X_2)$.

For a proof of this see Murray and von Neumann(1936), Lemma 2.2.1 or Courant and Hilbert (1937), Sec II.1.6. Lemma 4.1 shows that as in Lemma 2.1 and Corollary 2.1 decomposable elements of the form $\phi_i \otimes \psi_j = \phi_i \cdot \psi_j$ generate the whole $L^2(X_1, X_2)$ space.

Now let us take a look at the inner product. Let $\phi_1, \phi_2 \in L^2(X_1)$, $\psi_1, \psi_2 \in L^2(X_2)$. Then

$$
\begin{aligned}
(\phi_1 \otimes \psi_1, \phi_2 \otimes \psi_2) \\
= \int \phi_1(x_1)\psi_1(x_2)\phi_2(x_1)\psi_2(x_2)\mu_1(dx_1)\mu_2(dx_2) \\
= \int \phi_1(x_1)\phi_2(x_1)\mu_1(dx_1) \int \psi_1(x_2)\psi_2(x_2)\mu_2(dx_2) \\
= (\phi_1, \phi_2) \cdot (\psi_1, \psi_2).
\end{aligned}
$$

(4.5)

This is the same relation as in Proposition 2.2. We see that the inner product of $L^2(X_1, X_2)$ corresponds to the inner product introduced to $R^m \otimes R^n$ in Section 2. Therefore all orthogonality relations of Section 2 can be translated here. In particular Theorem 2.1 can be generalized as

Theorem 4.1. *Let U_i be a closed subspace in $L^2(X_i)$, $i = 1, \ldots, n$. Let $U_1 \otimes \cdots \otimes U_n$ be defined as a closed subspace generated by $\{\phi_1 \otimes \cdots \otimes \phi_n, \phi_i \in U_i, i = 1, \ldots, n\}$. Let $U_i^0 = U_i$, $U_i^1 = U_i^\perp$ for convenience. Then 2^n subspaces $\{\otimes_{i=1}^n U_i^{\epsilon_i}, \epsilon_i = 0, 1, i = 1, \ldots, n\}$ form a decomposition of $L^2(X_1, \ldots, X_n)$ into mutually orthogonal closed subspaces.*

Now let $F_i : L^2(X_i) \to L^2(X_i)$, $i = 1, 2$, be bounded linear transformations. We define a linear operator (easily seen to be bounded) $F_1 \otimes F_2 : L^2(X_1, X_2) \to L^2(X_1, X_2)$ by

(4.6)
$$
(F_1 \otimes F_2)(\phi_1 \otimes \phi_2) = F_1\phi_1 \otimes F_2\phi_2
$$

for decomposable elements and extend by linearity. See Remark 2.3. For a further justification of this see Murray and von Neumann(1936).

Next we consider orthogonal projections. Note that the definition of orthogonal projector in (2.17) is independent of the dimensionality. Therefore with the same proof for Theorem 2.2 we have

Theorem 4.2. *Let $U_i \subset L^2(X_i)$, $i = 1, \ldots, n$ be closed subspaces and P_{U_i} be the orthogonal projectors onto U_i in $L^2(X_i)$, $i = 1, \ldots, n$. Then the orthogonal projector onto $\otimes_{i=1}^n U_i \subset L^2(X_1, \ldots, X_n)$ is given by $P_{U_1} \otimes \cdots \otimes P_{U_n}$.*

Now let us come back to ANOVA type decomposition. Let $1_i(x_i) \equiv 1 \in L^2(X_i)$ and $U_i = \text{span}\{1_i\} \subset L^2(X_i)$. Let F_i be a linear transformation corresponding to taking the mean. For $\phi \in L^2(X_i)$

$$
(4.7) \qquad
\begin{aligned}
F_i\phi &= \int \phi(x_i)\mu_i(dx_i) \\
&= \mathcal{E}\phi = (\mathcal{E}\phi)1_i \in L^2(X_i).
\end{aligned}
$$

Then $F_i 1_i = 1_i$ and $F_i\phi = 0$ for $\phi \in L^2(X_i)$ such that $(1_i, \phi) = 0$. Therefore F_i is the orthogonal projector onto U_i and $P_{U_i} = F_i$. Denoting the identity map of $L^2(X_i)$ by I_i we have $P_{U_i^\perp} = I_i - F_i$. Now we define $\mathcal{L}_{i_1 \ldots i_k}$ by (3.1) and $P_{i_1 \ldots i_k}$ by (3.3) with I_i, F_i replacing $\overset{m_i}{\underset{i}{I}}, \underset{i}{F}$.

To see how $P_{i_1 \ldots i_k}$ behaves we fix complete orthonormal systems $\{\phi_1^i, \phi_2^i, \ldots\}$ of $L^2(X_i)$, $i = 1, \ldots, n$, such that $\phi_1^i = 1_i$, $i = 1, \ldots, n$. Note that $P_{U_i}\phi_1^i = \phi_1^i = 1$, $P_{U_i}\phi_j^i = 0$, for $j \geq 2$. Also $(I_i - P_{U_i})\phi_1^i = 0$, $(I_i - P_{U_i})\phi_j^i = \phi_j^i$ for $j \geq 2$. Using these relations we obtain

$$
(4.8) \qquad
\begin{aligned}
P_{i_1 \ldots i_k}\phi_{j_1}^1 \otimes \cdots \otimes \phi_{j_n}^n &= \phi_{j_1}^1 \otimes \cdots \otimes \phi_{j_n}^n \quad &&\text{if} \quad j_\ell \geq 2 \quad \text{for} \\
& && \ell \in \{i_1, \ldots, i_k\} \quad \text{and} \quad j_\ell = 1 \\
& && \text{for} \quad \ell \notin \{i_1, \ldots, i_k\}, \\
&= 0 \quad &&\text{otherwise.}
\end{aligned}
$$

Now consider $S \in L^2(X_1, \ldots, X_n)$. By Lemma 4.1 $\{\phi_{j_1}^1 \otimes \cdots \otimes \phi_{j_n}^n\}$ forms a basis of $L^2(X_1, \ldots, x_n)$. Hence we can write

$$
(4.9) \qquad S = \sum_{j_1} \cdots \sum_{j_n} a_{j_1 \ldots j_n} \phi_{j_1}^1 \otimes \cdots \otimes \phi_{j_n}^n,
$$

where

$$
a_{j_1 \ldots j_n} = \int S(x_1, \ldots, x_n)\phi_{j_1}^1 \cdots \phi_{j_n}^n \mu_1(dx_1) \cdots \mu_n(dx_n).
$$

Using (4.8) we obtain the following theorem.

Theorem 4.3.

$$
(4.10) \qquad P_{i_1 \ldots i_k}S = \sum_{j_\ell = 2, 1 \leq \ell \leq k}^{\infty} b_{j_1 \ldots j_k}^{i_1 \ldots i_k} \phi_{j_1}^{i_1} \cdots \phi_{j_k}^{i_k},
$$

where

$$
b_{j_1 \ldots j_k}^{i_1 \ldots i_k} = \int S(x_1, \ldots, x_n)\phi_{j_1}^{i_1}(x_{i_1}) \cdots \phi_{j_k}^{i_k}(x_{i_k})\mu_1(dx_1) \cdots \mu_n(dx_n).
$$

Remark 4.1. Actually (4.10) does not cover the case $k = 0$. In this case $P_\emptyset S = a_{1...1} = \mathcal{E} S$.

Theorem 4.1 gives a "coordinatewise" description of $P_{i_1...i_k}$ given complete orthonormal systems.

In Efron and Stein(1981), Bhargava(1980) and Karlin and Rinott(1982) these projections are given using conditional expectations. We will show that two definitions are the same. Let $E_i : L^2(X_1, \ldots, X_n) \to L^2(X_1, \ldots, X_n)$ be defined by

$$(4.11) \quad \begin{aligned} E_i \phi &= \int \phi(x_1, \ldots, x_n) \mu_i(dx_i) \\ &= \mathcal{E}(\phi \mid x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n). \end{aligned}$$

Let I denote the identity map in $L^2(X_1, \ldots, X_n)$. Then

$$(4.12) \quad (I - E_i)\phi = \phi - \mathcal{E}(\phi \mid x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n).$$

Let

$$(4.13) \quad H_{i_1...i_k} = G_1 \circ \cdots \circ G_n,$$

where $\circ$ denotes composition of the maps and

$$\begin{aligned} G_i &= I - E_i \quad if \quad i \in \{i_1, \ldots, i_k\}, \\ &= E_i \quad otherwise. \end{aligned}$$

Then

Theorem 4.4.

$$(4.14) \quad H_{i_1...i_k} = P_{i_1...i_k}.$$

Proof: Since $\{\phi^1_{j_1} \otimes \cdots \otimes \phi^n_{j_n}\}$ forms a basis it suffices to prove that

$$H_{i_1...i_k}(\phi^1_{j_1} \otimes \cdots \otimes \phi^n_{j_n}) = P_{i_1...i_n}(\phi^1_{j_1} \otimes \cdots \otimes \phi^n_{j_n})$$

for all $(j_1, \ldots, j_n)$. Let $\psi = \phi^1_{j_1} \otimes \cdots \otimes \phi^n_{j_n}$. Then

$$E_i \psi = \int \phi^1_{j_1} \otimes \cdots \otimes \phi^n_{j_n} \mu_i(dx_i)$$
$$= \phi^1_{j_1} \cdots (\mathcal{E}\phi^i_{j_i}) 1_i \cdots \phi^n_{j_n}(x_n).$$

Hence

$$E_i \psi = 0 \quad if \ j_i \geq 2$$
$$= \psi \quad if \ j_i = 1.$$

From this it follows that

$$
\begin{aligned}
H_{i_1 \ldots i_k}(\phi^1_{j_1} \otimes \cdots \otimes \phi^n_{j_n}) = \phi^1_{j_1} &\otimes \cdots \otimes \phi^n_{j_n} \quad if \quad j_\ell \geq 2 \quad for \\
&\ell \in \{i_1, \ldots, i_k\} \quad and \quad j_\ell = 1 \\
&for \quad \ell \notin \{i_1, \ldots, i_k\}, \\
= 0 \quad &otherwise.
\end{aligned}
$$

(4.15)

This is identical to (4.8). ∎

If we expand the right hand side of (4.13) we obtain the inclusion-exclusion pattern of conditional expectations. For example

$$
\begin{aligned}
H_1 S &= (I - E_1) \circ E_2 \circ \cdots \circ E_n S \\
&= \int S \mu_2(dx_2) \cdots \mu_n(dx_n) - \int S \mu_1(dx_1) \cdots \mu_n(dx_n) \\
&= \mathcal{E}(S \mid x_1) - \mathcal{E}(S).
\end{aligned}
$$

(4.16)

$$
\begin{aligned}
H_{12} S &= (I - E_1) \circ (I - E_2) \circ E_3 \circ \cdots \circ E_n S \\
&= E_3 \circ \cdots \circ E_n S - E_1 \circ E_3 \circ \cdots \circ E_n S \\
&\quad - E_2 \circ E_3 \circ \cdots \circ E_n S + E_1 \circ E_2 \circ \cdots \circ E_n S \\
&= \mathcal{E}(S \mid x_1, x_2) - \mathcal{E}(S \mid x_2) - \mathcal{E}(S \mid x_1) + \mathcal{E}(S).
\end{aligned}
$$

(4.17)

These expressions are used as definitions of the terms of ANOVA type decomposition in Efron and Stein(1981), Bhargava(1980), and Karlin and Rinott(1982).

References

Bhargava, R.P. (1980). A property of the Jackknife estimation of the variance when more than one observation is omitted. Tech. Report No.140, Dept. Statist., Stanford Univ.

Bock, R.D. (1963). Programming univariate and multivariate analysis of variance. *Technometrics*, **5**, 95-117.

Cotter, S.C. (1974). A general method of confounding for symmetrical factorial experiments. *J.Roy.Statist.Soc. B*, **36**, 267-276.

Cotter, S.C. (1975). Partial confounding in symmetrical factorial experiments. *J.Roy.Statist.Soc B*, **37**, 54-62.

Cotter, S.C., John, J.A., and Smith, T.M.F. (1973). Multi-factor experiments in non-orthogonal designs. *J.Roy.Statist.Soc B*, **35**, 361-367.

Courant, R. and Hilbert, D. *Methods of mathematical physics, Volume I.*Wiley, New York.

Dieudonné, J. (1976). *Treatise on analysis, Volume II.* Academic Press, New York.

Eaton, M.L. (1970). Gauss-Markov estimation for multivariate linear models: a coordinate free approach. *Ann.Math.Statist.*, **41**, 528-538.

Efron, B. and Stein, C. (1981). The Jackknife estimate of variance. *Ann.Statist.* **9**,586-596.

Federer, W.T. and Zelen, M. (1966). Analysis of multifactor classifications with unequal number of observations. *Biometrics*, **22**, 525-552.

Greub, W. (1978). *Multilinear Algebra, 2nd ed.* Springer-Verlag, New York.

Haberman, S.J. (1974). *The analysis of frequency data.* Univ. of Chicago Press.

Haberman, S.J. (1975). Direct products and linear models for complete factorial tables. *Ann.Statist.*, **3**, 314-333.

Hajek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann.Math.Statist.*, **39**, 325-346.

Han, T.S. (1977). Lattice-theoretic formulation of multifactor interactions and minimality of cell-symmetric decomposition of χ^2-statistics. *Rep.Statist.Appl.Res.Un.Japan.Sci.Engrs.*,

24, 55-66.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distributins. *Ann.Math. Statist.,* **19,** 293-325.

Jacobsen, R.L. (1968). Linear algebra and ANOVA. Thesis, Cornell University.

John, J.A. and Dean, A.M. (1975a). Single replicate factorial experiments in generalized cyclic designs: I. Symmetrical arrangements. *J.Roy.Statist.Soc. B,* **37,** 63-71.

John, J.A. and Dean, A.M. (1975b). Single replicate factorial experiments in generalized cyclic designs: I. Asymmetrical arrangements. *J.Roy.Statist.Soc. B,* **37,** 72-76.

Karlin, S. and Rinott, Y. (1982). Applications of ANOVA type decompositions for comparisons of conditional variance statistics including Jackknife estimates. *Ann.Statist.,* **10,** 485-501.

Kurkjian, B. and Zelen, M. (1962). A calculus for factorial arrangements. *Ann.Math.Statist.,* **33,** 600-619.

Kurkjian, B. and Zelen, M. (1963). Applications of the calculus for factorial arrangements. I. Block and direct product designs. *Biometrika,* **50,** 63-73.

Mann, H.B. (1949). *Analysis and design of experiments.* Dover, New York.

Maurin, K. (1967). *Methods of Hilbert spaces.* Polish Scientific Publishers, Warsaw.

Murray, F.J. and Neumann, J.v. (1936). On rings of operators. *Ann.Math.,* **37,** 116-229.

Paik, U.B. and Federer, W.T. (1974). Analysis of nonorthogonal *n*-way classifications. *Ann.Statist.,* **2,** 1000-1021.

Rubin, H. and Vitale, R.A. (1980). Asymptotic distributions of symmetric statistics. *Ann.Statist.,* **8,** 165-170.

Scheffé,H. (1959). *The analysis of variance.* Wiley, New York.

Serfling, R.J. (1980). *Approximation theorems of mathematical statistics.* Wiley, New York.

Sokolnikoff, I.S. (1964). *Tensor analysis, theory and application to geometry and mechanics of continua, 2nd ed.* Wiley, New York.

Zelen, M. and Federer, W.T. (1964). Applications of the calculus for factorial arrangements.

II. Designs with two-way elimination of heterogeneity. *Ann.Math.Statist.* **35**, 658-672.

Zelen, M. and Federer, W.T. (1965). Applications of the calculus for factorial arrangements. III. Analysis of factorials with unequal numbers of observations. *Sankhya A,* **27,** 383-400.

TECHNICAL REPORTS

U.S. ARMY RESEARCH OFFICE - CONTRACT DAAG29-82-K-0156

1. "Mamimum Likelihood Estimators and Likelihood Ratio Criteria for
   Multivariate Elliptically Contoured Distributions," T.W. Anderson and
   Kai-Tai Fang, September 1982.

2. "A Review and Some Extensions of Takemura's Generalizations of
   Cochran's Theorem," George P.H. Styan, September 1982.

3. "Some Further Applications of Finite Difference Operators," Kai-Tai Fang,
   September 1982.

4. "Rank Additivity and Matrix Polynomials," George P.H. Styan and
   Akimichi Takemura, September 1982.

5. "The Problem of Selecting a Given Number of Representative Points in a
   Normal Population and a Generalized Mills' Ratio," Kai-Tai Fang and
   Shu-Dong He, October 1982.

6. "Tensor Analysis of ANOVA Decomposition," Akimichi Takemura, November 1982.

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 6	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Tensor Analysis of ANOVA Decomposition		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR*(s)* Akimichi Takemura		8. CONTRACT OR GRANT NUMBER*(s)* DAAG29-82-K-0156
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics - Sequoia Hall Stanford University Stanford, California 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P-19065-M
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office		12. REPORT DATE November 1982
		13. NUMBER OF PAGES 18
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

ANOVA, crossed layout, ANOVA-type decomposition, tensor product, direct product, Kronecker product, U-statistics.

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

The analogy between ANOVA for general n-factor crossed layouts and the ANOVA-type decomposition of square integrable statistics is demonstrated using the notions and notations of tensor analysis and multilinear algebra. A theory of tensors is developed in such a way that (i) it can be immediately applied in computer programs, (ii) it can be easily generalized to L^2 spaces.

DD $_{1\ JAN\ 73}^{FORM}$ 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601 !